DIFFUSIONNET: AN EFFICIENT FRAMEWORK TO CLASSIFY SINGLE-MOLECULE IMAGES WITH LATENT ENTROPY MINIMIZATION

Soumee Guha[1]*, Olivia de Cuba*[2]*, Andreas Gahlmann*[2,3]*, Scott T. Acton*[1]

[1]Department of Electrical and Computer Engineering, University of Virginia, VA, USA
[2]Department of Molecular Physiology
Biological Physics, University of Virginia School of Medicine, VA, USA
[3]Department of Chemistry, University of Virginia, VA, USA

ABSTRACT

Single-molecule tracking is a powerful tool to measure the dynamics of proteins in living cells. Analysis of molecule diffusion aids our understanding of molecular mechanisms and helps us distinguish between freely diffusing cytosolic protein populations and the more slowly moving membrane-associated population. There is no existing technique that distinguishes the diffusion coefficients of single-molecule images without analyzing their trajectories, which extend over multiple sequential camera frames. To this end, we propose a spatial and channel attention-based convolutional neural network (CNN) architecture with latent entropy minimization that efficiently classifies individual single-molecule images by the imaged molecules' diffusion coefficients. We also propose a loss function which minimizes the entropy of the attention maps. Experiments demonstrate that the diffusion coefficients can efficiently identify different types of molecules in simulated and experimental datasets and our method outperforms state-of-the-art models for image classification (e.g., Resnet-18 and Densenet).

***Index Terms*—** Single-molecule tracking, diffusion coefficient, protein, attention, latent entropy minimization

1. INTRODUCTION

Single-molecule localization microscopy (SMLM) describes a group of imaging modalities with improved spatial resolution over conventional, diffraction-limited techniques, making it exceptionally well suited to image biological structures and study processes at the molecular scale. The ability to express fluorescent fusion proteins in bacterial cells, i.e. a protein of interest directly fused with/labeled by a fluorescent protein, has opened the door to spatiotemporal studies of proteins to determine how and when protein-protein interactions may be involved in larger cellular processes. In conventional single-molecule tracking, the repeated detection and localization of a given molecule over time generates trajectories that can be analyzed for the mode of motion, e.g. directed motion or random (Brownian) diffusion. For Brownian motion, the diffusion coefficient is a fundamental characteristic, as it depends on the molecules size and possibly its shape. As such, the binding of a protein to another biomolecule is coupled to a change in diffusive state. For example, we can track a protein that can exist in two distinct diffusive states: it has one state that is freely diffusing in the cytosol and a state that is membrane- or DNA-bound. While there are a range of approaches to single-molecule tracking [1, 2, 3, 4, 5, 6], including some that apply a convolutional neural network (CNN) for the classification of the trajectories, to our knowledge there is no work that aims to classify individual single-molecule images by the imaged molecules' diffusion coefficient. Such an approach would be particularly useful for identifying changes in diffusive behavior, such as binding and unbinding events of a fluorescent-labeled protein during a single trajectory. Common approaches, such as mean squared displacement (MSD) analysis, average over changes in diffusion making it difficult to detect these biologically relevant events. Being able to identify such changes in diffusion is thus critical to generate biological insights and needs to be automated.

In this work, we show that it is possible to perform binary classification on individual single-molecule images in which the molecules have different diffusion coefficients. We test our binary classification approach on images of dye-labeled ($SspB_{micro}$) proteins that can bind to a membrane-bound interacting partner (iLID). We have previously determined this protein's cytosolic diffusive state to be 5.5 $\mu m^2/s$ while the membrane-associated state is 0.2 $\mu m^2/s$ [7]. Through cumulative displacement analysis of (multi-frame) trajectory segments, we have manually identified a subset of trajectories that display state switching behavior (i.e. trajectories that contain fast *and* slow segments). Herein, we provide a strong alternative to manual annotation to identify state-switching events in single-molecule trajectories through single image analysis. As a preliminary work, we consider integer-valued diffusion coefficients 1 through 10 $\mu m^2/s$, an appropriate range for bacterial fusion proteins that may be cytosolic or

This work was funded in part by NIH 1R01 GM139002.

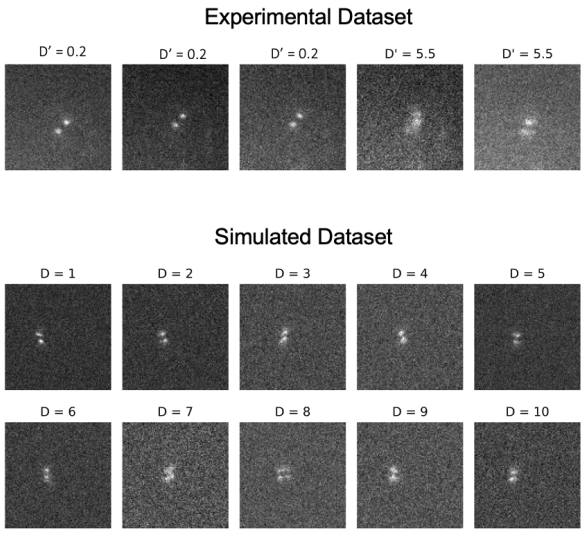

Fig. 1. Row 1: Sample microscope images from the experimental dataset. D' is the estimated diffusion coefficient for each of the images. Rows 2 and 3: Sample images from the simulated dataset for each of the 10 diffusion coefficients (D).

membrane-associated, as reviewed elsewhere [8].

The SMLM images used for this study are typically low in signal-to-noise ratio (SNR) and the images of different classes are very similar to each other due the combination of motion blurring and confinement of the proteins inside the small bacterial cells (Figure 1). We propose a CNN architecture that possesses spatial and channel attention blocks with a minimum entropy constraint imposed directly on the attention maps. Unlike [9], the spatial attention block used in this work is a sequence of convolutional filtering and activation function and does not involve non-maximum suppression and ranking the attention map patches. Existing studies on minimum entropy latent models include optimal latent parameter estimation [10], minimizing the entropy of the output distribution [11, 12] and minimizing the differential of the latent space entropy in an autoregressive framework [13]. In contrast, we introduce an entropy minimization term on the learned spatial attention maps directly.

The main contributions of this paper are: (1) We are the first to present a CNN architecture (DiffusionNet) to efficiently identify state-switching events by classifying single-molecule images based on the molecules' diffusion coefficients. (2) We show that entropy minimization of learned attention maps, when combined with cross-entropy loss results in better accuracy and stable classification scores across all class labels as compared to using cross-entropy loss only. (3) We show that the knowledge of diffusion coefficients can be used to distinguish between different types of (macromolecular) states from SMLM images. The data and code are available at https://github.com/soumeeguha/DiffusionNet.

2. DATASETS

Experimental Dataset: Experimental images of K12 Escherichia coli (MG1655) expressing the membrane-anchored, improved Light Induced Dimerization (iLID) system and its interacting partner $SspB_{micro}$-Halo, stained with JFX-549 dye, was induced and imaged as described elsewhere [7, 14]. In brief, overnight *E. coli* cultures were diluted and the $SspB_{micro}$-Halo proteins were stained with JFX-549 dye. Cells were imaged on a home-built inverted fluorescence microscope. Emitted fluorescence within a spectral window of 570 nm – 700 nm was captured using an sCMOS camera (Hamamatsu, ORCA Flash 4.0 v2). The iLid system was not activated (by λ_{488}) for this sample dataset.

Simulated Dataset: Because generating experimentally obtained paired data is an inefficient process, we generated a simulated dataset as described [15] for the purposes of obtaining a sufficient training set. Briefly, we performed Monte Carlo simulations of confined Brownian motion for diffusion coefficients ranging from 1–10 $\mu m^2/s$ within a volume matching that of a bacterial cell. Motion-blurred single-molecule images using the double-helix point spread function (DH-PSF) were generated from simulated Brownian Motion trajectories. The DH-PSF comprises two lobes which rotate along the axial direction, thereby encoding both lateral and axial/depth information in a 2D image [16, 17]. For each trajectory, 50 positions of a molecule during the camera exposure time (25 ms) were sampled and these 50 sub-images were summed to obtain the motion-blurred DH-PSF images. The photon count of each simulated image was scaled to match the experimentally measured distribution for eYFP photon counts and a laser background, Poisson noise and dark offset were added to produce the final image.

3. PROPOSED DIFFUSIONNET

For 10 sets of single molecule images corresponding to the 10 diffusion coefficients (1–10 $\mu m^2/s$), given any two different sets of images, DiffusionNet classifies them according to their diffusion coefficients.

Preprocessing: We apply non-local means denoising to all the images, which improves the SNR without compromising the finer details of the images [18, 19]. The molecule is identified in each image and a patch containing the molecule is cropped out using a filtering method. We concatenate the magnitudes of the Wavelet Transform coefficients along with the magnitude spectrum of Fourier coefficients in order to identify the global and local characteristics of the images.

Double Attention Block: Inspired by [9], we have introduced spatial and channel attention separately that helps in localizing the discriminating regions along the spatial dimension and along the channels sequentially. Since the im-

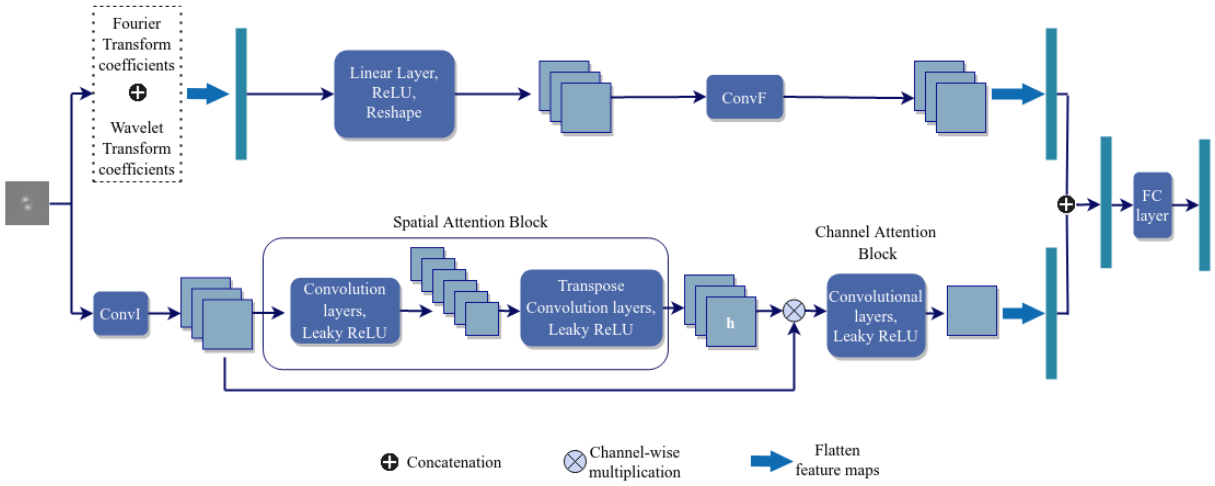

Fig. 2. The DiffusionNet architecture

Diffusion Coefficient (D) ($\mu m^2/s$)	1	2	3	4	5	6	7
4	95.8	-	-	-	-	-	-
5	95.3	71.5	-	-	-	-	-
6	98.1	81.5	71.2	-	-	-	-
7	-	84.9	76.5	70.8	-	-	-
8	-	-	84.6	82.2	73.9	-	-
9	-	-	-	87.1	77.1	64.8	-
10	-	-	-	-	88.2	84.9	80.1

Table 1. DiffusionNet performance for the different diffusion coefficient pairs. The (i, j) element refers to the prediction accuracy (%) for images having diffusion coefficient i and diffusion coefficient j.

Method	$\Delta D = 3$	$\Delta D = 4$	$\Delta D = 5$
DiffusionNet w/o Spatial and Channel Attention	72.01	74.85	80.3
DiffusionNet w/o Spatial Attention	72.79	74.86	81.72
DiffusionNet	71.45	74.94	81.78
DiffusionNet + Latent Entropy Minimization	**75.43**	**82.92**	**88.55**

Table 2. Comparative performance (accuracy values in %) on the simulated dataset

ages are very similar, downsampling the feature maps in the spatial attention block condenses the discriminating information. Intuitively, upsampling the condensed information helps the model identify the subtle differences in a more refined manner. Additionally, we learn the relationship between the different channels of the feature maps separately as shown in Figure 2. Unlike our method, [9] has implemented the spatial attention as a pooling operation and has a applied non-maximum suppression and a ranking function on the learned attention maps to better identify the discriminating regions.

Latent Variable Entropy Minimization: Let $X = \{x_i\}_{i=1}^m$ and $Y = \{y_i\}_{i=1}^m$ represent the input images and their corresponding labels respectively where m is the cardinality of the training set. Let h_i denote the latent variable generated by the spatial attention block for input x_i (Figure 2) and θ denote the model parameters. The optimal model is given by $\theta^* = \text{argmax}_\theta\, P(Y, h|X; \theta)$, where h represents the latent variables [10]. Assuming that the images are independent and identically distributed,

$$\theta^* = \text{argmax}_\theta\, P(Y, h|X; \theta)$$

$$= \text{argmax}_\theta\, P(h|Y, X; \theta)P(Y|X; \theta)$$

$$= \text{argmax}_\theta \sum_{i=1}^m logP(h_i|y_i, x_i; \theta) + \sum_{i=1}^m logP(y_i|x_i; \theta)$$

$$= \text{argmax}_\theta \sum_{i=1}^m \tfrac{1}{m} logP(h_i|y_i, x_i; \theta) + \sum_{i=1}^m \tfrac{1}{m} logP(y_i|x_i; \theta)$$

$$= \text{argmax}_\theta\, E_{h\sim P}\, log\, P(h|Y, X; \theta) + E_{x\sim P^*}\, log\, P(Y|X; \theta)$$

$$= \text{argmin}_\theta\, LE + CE$$

Here, P^* and P represent the true data distribution and

Method	Accuracy
Leave-one-out with augmented real images	96.58
Simulated images + augmented 10% real images	98.48

Table 3. Performance on experimental dataset (SMLM images)

the predicted distribution respectively. LE and CE denote the entropy of the latent variables and the cross entropy loss respectively. Hence, our loss function is $\alpha LE + CE$, where α is a hyperparameter.

4. EXPERIMENTS

4.1. Results

Simulated Dataset: Each class contained 600 images. The images were 30 x 30 in size after denoising and cropping the region of interest. The training converged in 20 epochs. We used the Adam optimizer [20] with an initial learning rate of 0.0001 and after 10 epochs, it was reduced to 0.00005. We used the loss function defined in Section 3. In Table 1, we report the average accuracy values for 10 experiments where every time the training and validation sets was selected randomly. For each of the individual cases, the F1 score varied within 6% of the accuracy scores mentioned in Table 1. Table 2 summarizes the performance of our model and compares it to the other variations of the model. We used Resnet-18 and Densenet for comparison. The test accuracies for both the models were 50% for all the individual cases. This poor performance is due to the fact that the training images are 30x30 in size and all the classes are very similar to one another. ΔD refers to the difference in diffusion coefficients between the two sets of images. We see that DiffusionNet with latent variable entropy minimization performs the best.

Experimental Dataset: We used the proposed DiffusionNet with the loss function proposed in Section 3 to train and test the model with SMLM images (experimental dataset). As described previously, we have two molecules diffusing approximately at 0.2 $\mu m^2/s$ and 5.5 $\mu m^2/s$. In the first experiment, we randomly separate the dataset into train and test sets, where each test set contains one image of each class (leave-one-out). We report the average test accuracy achieved by training the network on augmentations of the training images in Table 3. In the second experiment, we randomly select 10% of the experimental images (same number of images for both the classes) and train the model with their augmentations and simulated images. We used the (1, 5) and (1, 6) models separately and the final performance was similar in both the cases. The average test accuracy is reported in Table 3.

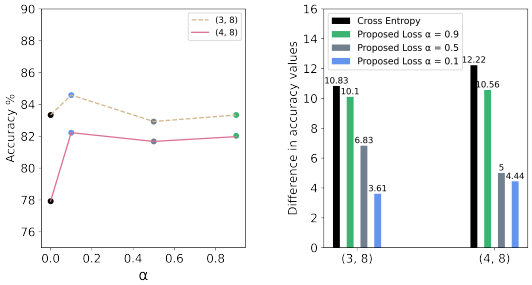

Fig. 3. Left: The effect of α on predicted accuracy values. Right: The effect of α on difference between accuracy values of the two classes (where lower values are better).

4.2. Ablation Study

We study the effectiveness of the latent entropy minimization term in our loss function with two different sets of images, (3, 8) and (4, 8). We trained the proposed DiffusionNet with Cross-Entropy loss and the proposed loss separately. In Figure 3, the first plot shows that for the Cross-Entropy loss and the proposed loss with variation in α and the maximum accuracy for both the cases occur at $\alpha = 0.1$. The second plot shows the difference in the prediction accuracies of the two classes in these two sets of images, (3, 8) and (4, 8). Although there is not much difference in the average accuracies for the different values of α, the difference in the accuracy values for the individual classes is minimum when $\alpha = 0.1$. The difference is maximum when only Cross-Entropy loss is used. A smaller difference is desired as the average accuracy is a better representation for the individual class accuracies.

5. CONCLUSION AND FUTURE WORK

In this work, we present an efficient architecture for classifying SMLM images based on the diffusion coefficients of the imaged single molecules. We also showed that using a latent entropy minimization term along with the Cross-Entropy loss leads to better performance. Low SNR, high intra-class variance, and low inter-class variance make the classification task complicated. The performance of our model is similar for the simulated dataset and the SMLM images. Our approach can be used to pinpoint protein binding and unbinding events from SMLM experiments in living cells in an automated pipeline. However, this requires some *a priori* knowledge of the true number of diffusive states for a protein of interest. Next, we will extend our work to multiclass prediction which would enable this analysis to be applied to more complex, multi-state biological systems.

6. REFERENCES

[1] JJ Maris, Freddy T Rabouw, Bert M Weckhuysen, and Florian Meirer, "Classification-based motion analysis of single-molecule trajectories using diffusionlab," *Scientific Reports*, vol. 12, no. 1, pp. 1–10, 2022.

[2] Xavier Michalet and Andrew J Berglund, "Optimal diffusion coefficient estimation in single-particle tracking," *Physical Review E*, vol. 85, no. 6, pp. 061916, 2012.

[3] Andrew J Berglund, "Statistics of camera-based single-particle tracking," *Physical Review E*, vol. 82, no. 1, pp. 011917, 2010.

[4] Dominique Ernst and Jürgen Köhler, "Measuring a diffusion coefficient by single-particle tracking: statistical analysis of experimental mean squared displacement curves," *Physical Chemistry Chemical Physics*, vol. 15, no. 3, pp. 845–849, 2013.

[5] Michael J Saxton, "Single-particle tracking: the distribution of diffusion coefficients," *Biophysical journal*, vol. 72, no. 4, pp. 1744–1753, 1997.

[6] Naor Granik, Lucien E Weiss, Elias Nehme, Maayan Levin, Michael Chein, Eran Perlson, Yael Roichman, and Yoav Shechtman, "Single-particle diffusion characterization by deep learning," *Biophysical journal*, vol. 117, no. 2, pp. 185–192, 2019.

[7] Alecia Marie Achimovich, Ting Yan, and Andreas Gahlmann, "Dimerization of ilid optogenetic proteins observed using 3d single-molecule tracking in live bacterial cells," *bioRxiv*, 2022.

[8] Paul E Schavemaker, Arnold J Boersma, and Bert Poolman, "How important is protein diffusion in prokaryotes?," *Frontiers in Molecular Biosciences*, vol. 5, pp. 93, 2018.

[9] Chao Liu, Lei Huang, Zhiqiang Wei, and Wenfeng Zhang, "Subtler mixed attention network on fine-grained image classification," *Applied Intelligence*, vol. 51, no. 11, pp. 7903–7916, 2021.

[10] Kevin Miller, M Pawan Kumar, Ben Packer, Danny Goodman, and Daphne Koller, "Max-margin min-entropy models," in *Artificial Intelligence and Statistics*. PMLR, 2012, pp. 779–787.

[11] Fang Wan, Pengxu Wei, Jianbin Jiao, Zhenjun Han, and Qixiang Ye, "Min-entropy latent model for weakly supervised object detection," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2018, pp. 1297–1306.

[12] Diane Bouchacourt, Sebastian Nowozin, and M Pawan Kumar, "Entropy-based latent structured output prediction," in *Proceedings of the IEEE International Conference on Computer Vision*, 2015, pp. 2920–2928.

[13] Davide Abati, Angelo Porrello, Simone Calderara, and Rita Cucchiara, "Latent space autoregression for novelty detection," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2019, pp. 481–490.

[14] Gurkan Guntas, Ryan A. Hallett, Seth P. Zimmerman, Tishan Williams, Hayretin Yumerefendi, James E. Bear, and Brian Kuhlman, "Engineering an improved light-induced dimer (ilid) for controlling the localization and activity of signaling proteins," *Proceedings of the National Academy of Sciences*, vol. 112, no. 1, pp. 112–117, 2015.

[15] Julian Michael Rocha, Charles Joseph Richardson, Mingxing Zhang, Caroline Maureen Darch, Eugene Cai, Andreas Diepold, and Andreas Gahlmann, "Single-molecule tracking in live yersinia enterocolitica reveals distinct cytosolic complexes of injectisome subunits," *Integrative Biology*, vol. 10, no. 9, pp. 502–515, 2018.

[16] Jorge Zepeda O, Logan D. C. Bishop, Chayan Dutta, Suparna Sarkar-Banerjee, Wesley W. Leung, and Christy F. Landes, "Untying the gordian knot: Unbiased single particle tracking using point clouds and adaptive motion analysis," *The Journal of Physical Chemistry A*, vol. 125, no. 39, pp. 8723–8733, 2021, PMID: 34559965.

[17] Sri Rama Prasanna Pavani, Michael A. Thompson, Julie S. Biteen, Samuel J. Lord, Na Liu, Robert J. Twieg, Rafael Piestun, and W. E. Moerner, "Three-dimensional, single-molecule fluorescence imaging beyond the diffraction limit by using a double-helix point spread function," *Proceedings of the National Academy of Sciences*, vol. 106, no. 9, pp. 2995–2999, 2009.

[18] Antoni Buades, Bartomeu Coll, and J-M Morel, "A non-local algorithm for image denoising," in *2005 IEEE computer society conference on computer vision and pattern recognition (CVPR'05)*. Ieee, 2005, vol. 2, pp. 60–65.

[19] Antoni Buades, Bartomeu Coll, and Jean-Michel Morel, "Non-local means denoising," *Image Processing On Line*, vol. 1, pp. 208–212, 2011.

[20] Diederik P Kingma and Jimmy Ba, "Adam: A method for stochastic optimization," *arXiv preprint arXiv:1412.6980*, 2014.